ECON 351* -- NOTE 23

Tests for Coefficient Differences: Examples 2

1. Introduction

Model and Data

- **Sample data:** A random sample of 534 paid employees.

- **Variable definitions:**

  W_i = hourly wage rate of employee i;
  $\ln W_i$ = the natural logarithm of W_i;
  S_i = years of schooling completed by employee i;
  X_i = years of work experience accumulated by employee i.
  F_i = a female indicator variable, = 1 if employee i is female, 0 otherwise;
  M_i = a male indicator variable, = 1 if employee i is male, 0 otherwise.

- **The Model:** A simple log-lin (semi-log) wage equation of the form

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + u_i \qquad \textbf{(1)}$$

- **Two Groups of Employees: Female and Male**

♦ The ***female*** **wage equation**

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + u_i \qquad i = 1, ..., N_f = 245 \qquad \textbf{(2.1)}$$

♦ The ***male*** **wage equation**

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + u_i \qquad i = 1, ..., N_m = 289 \qquad \textbf{(2.2)}$$

2. Tests for Full Coefficient Equality

Null and Alternative Hypotheses

H_0: $\alpha_0 = \beta_0$ *and* $\alpha_1 = \beta_1$ *and* $\alpha_2 = \beta_2$ (3 coefficient restrictions)

H_1: $\alpha_0 \neq \beta_0$ *and/or* $\alpha_1 \neq \beta_1$ *and/or* $\alpha_2 \neq \beta_2$.

***Unrestricted* Model – Approach 1: *Separate* Female and Male Wage Equations**

♦ The ***female* wage equation**

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + u_i \qquad i = 1, ..., N_f = 245 \qquad \textbf{(2.1)}$$

The **OLS sample regression equation for *females*** (with t-ratios) is:

$$\ln W_i = \hat{\alpha}_0 + \hat{\alpha}_1 S_i + \hat{\alpha}_2 X_i + \hat{u}_i \qquad i = 1, ..., N_f = 245 \qquad K_0 = 3$$

$\hat{\alpha}_0 = 0.3031$ (1.699) $\quad \hat{\alpha}_1 = 0.1117$ (9.381) $\quad \hat{\alpha}_2 = 0.008854$ (3.868) $\quad$ $\mathbf{RSS_{(1)} = 43.2866}$ $\quad$ $\mathbf{df_{(1)} = 245 - 3 = 242}$

♦ The ***male* wage equation**

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + u_i \qquad i = 1, ..., N_m = 289 \qquad \textbf{(2.2)}$$

The **OLS sample regression equation for *males*** (with t-ratios) is:

$$\ln W_i = \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 X_i + \hat{u}_i \qquad i = 1, ..., N_m = 289 \qquad K_0 = 3$$

$\hat{\beta}_0 = 0.6965$ (4.279) $\quad \hat{\beta}_1 = 0.09030$ (8.430) $\quad \hat{\beta}_2 = 0.01635$ (6.690) $\quad$ $\mathbf{RSS_{(2)} = 63.1474}$ $\quad$ $\mathbf{df_{(2)} = 289 - 3 = 286}$

♦ The ***unrestricted* residual sum-of-squares** is:

$\mathbf{RSS_1} = RSS_{(1)} + RSS_{(2)} = 43.2866 + 63.1474 = \mathbf{106.4340}$;

$\mathbf{df_1} = df_{(1)} + df_{(2)} = 242 + 286 = \mathbf{528}$.

Unrestricted **Model – Approach 2: *Pooled* Full-Interaction Wage Equations**

- ***No* base group** – no intercept coefficient (t-ratios in parentheses)

$$\ln W_i = \alpha_0 F_i + \alpha_1 F_i S_i + \alpha_2 F_i X_i + \beta_0 M_i + \beta_1 M_i S_i + \beta_2 M_i X_i + u_i \qquad \textbf{(12.0)}$$

$$i = 1, ..., N = N_f + N_m = 534$$

$\mathbf{RSS_1 = 106.4340}$ *with* $\mathbf{df_1 = N - 2K_0 = 534 - 6 = 528}$

$\hat{\alpha}_0 = 0.3031$ (1.601) $\hat{\alpha}_1 = 0.1117$ (8.837) $\hat{\alpha}_2 = 0.008854$ (3.643)

$\hat{\beta}_0 = 0.6965$ (4.478) $\hat{\beta}_1 = 0.09030$ (8.822) $\hat{\beta}_2 = 0.01635$ (7.002)

$\hat{\alpha}_0 - \hat{\beta}_0 = -0.3933$ (– 1.605) $\hat{\alpha}_1 - \hat{\beta}_1 = 0.02144$ (1.318) $\hat{\alpha}_2 - \hat{\beta}_2 = -0.007491$ (–2.223)

$\hat{\beta}_0 - \hat{\alpha}_0 = 0.3933$ (1.605) $\hat{\beta}_1 - \hat{\alpha}_1 = -0.02144$ (– 1.318) $\hat{\beta}_2 - \hat{\alpha}_2 = 0.007491$ (2.223)

Compare coefficient estimates of pooled regression equation (12.0) with the ***separate*** female and male sample wage equations:

- The **OLS sample regression equation for *females*** is:

$$\ln W_i = \hat{\alpha}_0 + \hat{\alpha}_1 S_i + \hat{\alpha}_2 X_i + \hat{u}_i \qquad i = 1, ..., N_f = 245 \qquad K_0 = 3$$

$\hat{\alpha}_0 = 0.3031$ (1.699) $\hat{\alpha}_1 = 0.1117$ (9.381) $\hat{\alpha}_2 = 0.008854$ (3.868) $RSS_{(1)} = 43.2866$ $df_{(1)} = 245 - 3 = 242$

- The **OLS sample regression equation for *males*** is:

$$\ln W_i = \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 X_i + \hat{u}_i \qquad i = 1, ..., N_m = 289 \qquad K_0 = 3$$

$\hat{\beta}_0 = 0.6965$ (4.279) $\hat{\beta}_1 = 0.09030$ (8.430) $\hat{\beta}_2 = 0.01635$ (6.690) $RSS_{(2)} = 63.1474$ $df_{(2)} = 289 - 3 = 286$

♦ ***Females* as base group** (t-ratios in parentheses)

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + \gamma_0 M_i + \gamma_1 M_i S_i + \gamma_2 M_i X_i + u_i \qquad \textbf{(12.1)}$$

where $\gamma_0 = \beta_0 - \alpha_0$; $\gamma_1 = \beta_1 - \alpha_1$; $\gamma_2 = \beta_2 - \alpha_2$.

RSS$_1$ = 106.4340 *with* **df$_1$ = N – 2K$_0$ = 534 – 6 = 528**

$\hat{\alpha}_0 = 0.3031$ (1.601) $\hat{\alpha}_1 = 0.1117$ (8.837) $\hat{\alpha}_2 = 0.008854$ (3.643)

$\hat{\gamma}_0 = 0.3933$ (1.605) $\hat{\gamma}_1 = -0.02144$ (–1.318) $\hat{\gamma}_2 = 0.007491$ (2.223)

<u>From OLS-SRE (12.0)</u>:

$\hat{\beta}_0 - \hat{\alpha}_0 = 0.3933$ (1.605) $\hat{\beta}_1 - \hat{\alpha}_1 = -0.02144$ (–1.318) $\hat{\beta}_2 - \hat{\alpha}_2 = 0.007491$ (2.223)

♦ ***Males* as base group** (t-ratios in parentheses)

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \delta_0 F_i + \delta_1 F_i S_i + \delta_2 F_i X_i + u_i \qquad \textbf{(12.2)}$$

where $\delta_0 = \alpha_0 - \beta_0$; $\delta_1 = \alpha_1 - \beta_1$; $\delta_2 = \alpha_2 - \beta_2$.

RSS$_1$ = 106.4340 *with* **df$_1$ = N – 2K$_0$ = 534 – 6 = 528**

$\hat{\beta}_0 = 0.6965$ (4.478) $\hat{\beta}_1 = 0.09030$ (8.822) $\hat{\beta}_2 = 0.01635$ (7.002)

$\hat{\delta}_0 = -0.3933$ (–1.605) $\hat{\delta}_1 = 0.02144$ (1.318) $\hat{\delta}_2 = -0.007491$ (–2.223)

<u>From OLS-SRE (12.0)</u>:

$\hat{\alpha}_0 - \hat{\beta}_0 = -0.3933$ (–1.605) $\hat{\alpha}_1 - \hat{\beta}_1 = 0.02144$ (1.318) $\hat{\alpha}_2 - \hat{\beta}_2 = -0.007491$ (–2.223)

***Restricted* Model – Same for Approach 1 and Approach 2**

Corresponds to the null hypothesis that all female and male coefficients are equal:

H_0: $\alpha_0 = \beta_0$ *and* $\alpha_1 = \beta_1$ *and* $\alpha_2 = \beta_2$ in pooled equation (12.0)

$\gamma_0 = 0$ *and* $\gamma_1 = 0$ *and* $\gamma_2 = 0$ in pooled equation (12.1)

$\delta_0 = 0$ *and* $\delta_1 = 0$ *and* $\delta_2 = 0$ in pooled equation (12.2)

- The ***restricted* model** can be written as *either*

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + u_i \qquad i = 1, ..., N = 534 \qquad \textbf{(1.1)}$$

or

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + u_i \qquad i = 1, ..., N = 534 \qquad \textbf{(1.2)}$$

- The ***restricted* OLS-SRE** for the full sample of 534 observations is

$$\ln W_i = \tilde{\beta}_0 + \tilde{\beta}_1 S_i + \tilde{\beta}_2 X_i + \tilde{u}_i, \qquad i = 1, ..., N = N_1 + N_2 = 534$$

$\tilde{\beta}_0 = 0.5828$ (4.646) $\qquad \tilde{\beta}_1 = 0.09642$ (11.601) $\qquad \tilde{\beta}_2 = 0.01175$ (6.700)

- The ***restricted* residual sum-of-squares** is:

$\mathbf{RSS_0 = 117.0626}$ *with* $\mathbf{df_0 = N - K_0 = 534 - 3 = 531}$.

The F-Test for Equality of *All* Coefficients Between Males and Females

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N-K)} \sim F[df_0 - df_1, df_1]$$

$\mathbf{RSS_0 = 117.0626}$ *with* $\mathbf{df_0 = N - K_0 = 534 - 3 = 531}$
$\mathbf{RSS_1 = 106.4340}$ *with* $\mathbf{df_1 = N - 2K_0 = 534 - 6 = 528}$

- ***Sample value* of the F-statistic:**

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$
$$= \frac{(117.0626 - 106.4340)/(531 - 528)}{106.4340/528} \quad \Rightarrow \quad \mathbf{F_0 = 17.58}$$
$$= \frac{10.6286/3}{106.4340/528}$$
$$= 17.58$$

- ***Null distribution* of $\mathbf{F_0}$:** $F_0 \sim F[K_0, N - 2K_0]$ = F[3, 528] under H_0.

- ***Decision Rule:*** At the 100α percent significance level
  1. ***reject* $\mathbf{H_0}$** if $F_0 \geq F_\alpha[K_0, N - 2K_0]$ *or* p-value for $F_0 \leq \alpha$;
  2. ***retain* $\mathbf{H_0}$** if $F_0 < F_\alpha[K_0, N - 2K_0]$ *or* p-value for $F_0 > \alpha$.

- ***Critical Values of F[3, 528]:*** at the 5% and 1% significance levels.

  At α = 0.05: $F_{0.05}[3, 528] = 2.622$
  At α = 0.01: $F_{0.01}[3, 528] = 3.819$

- ***P-value for F_0* = 0.0000**.

- ***Inference:***

  Since $\mathbf{F_0 = 17.58 > 3.819 = F_{0.01}[3, 528]}$, ***reject*** $\mathbf{H_0}$ at the **1%** significance level.

  Since **p-value for $\mathbf{F_0}$ = 0.0000 < 0.01**, ***reject*** $\mathbf{H_0}$ at the **1%** significance level.

3. Tests for Equality of Both Slope Coefficients

Null and Alternative Hypotheses

female S_i coefficient = male S_i coefficient $\Rightarrow$ $\boldsymbol{\alpha_1 = \beta_1}$

female X_i coefficient = male X_i coefficient $\Rightarrow$ $\boldsymbol{\alpha_2 = \beta_2}$

H_0: $\alpha_1 = \beta_1$ *and* $\alpha_2 = \beta_2$ in pooled equation (12.0)
$\gamma_1 = 0$ *and* $\gamma_2 = 0$ in pooled equation (12.1)
$\delta_1 = 0$ *and* $\delta_2 = 0$ in pooled equation (12.2)

against

H_1: $\alpha_1 \neq \beta_1$ *and/or* $\alpha_2 \neq \beta_2$ in pooled equation (12.0)
$\gamma_1 \neq 0$ *and/or* $\gamma_2 \neq 0$ in pooled equation (12.1)
$\delta_1 \neq 0$ *and/or* $\delta_2 \neq 0$ in pooled equation (12.2)

***Unrestricted* Model – corresponds to H_1**

Any one of the three pooled full-interaction regression equations for $\ln W_i$.

$$\ln W_i = \alpha_0 F_i + \alpha_1 F_i S_i + \alpha_2 F_i X_i + \beta_0 M_i + \beta_1 M_i S_i + \beta_2 M_i X_i + u_i \qquad \textbf{(12.0)}$$

$$i = 1, ..., N = N_f + N_m = 534$$

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + \gamma_0 M_i + \gamma_1 M_i S_i + \gamma_2 M_i X_i + u_i \qquad \textbf{(12.1)}$$

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \delta_0 F_i + \delta_1 F_i S_i + \delta_2 F_i X_i + u_i \qquad \textbf{(12.2)}$$

- OLS estimation of any one of the three unrestricted regression equations (12.0), (12.1) or (12.2) on the pooled sample of 534 male and female employees yields the following value for the ***unrestricted*** **RSS**:

  $\mathbf{RSS_1 = 106.4340}$ *with* $\mathbf{df_1 = N - 2K_0 = 534 - 6 = 528}$.

Restricted **Model – corresponds to H_0**

Obtained by substituting the two coefficient restrictions specified by the null hypothesis H_0 into any one of the three pooled full-interaction regression equations for $\ln W_i$.

- **In equation (12.0) – *no base group*: set $\alpha_1 = \beta_1$ *and* $\alpha_2 = \beta_2$**

$$\ln W_i = \alpha_0 F_i + \alpha_1 F_i S_i + \alpha_2 F_i X_i + \beta_0 M_i + \beta_1 M_i S_i + \beta_2 M_i X_i + u_i \qquad (12.0)$$

$$\begin{aligned} \ln W_i &= \alpha_0 F_i + \beta_1 F_i S_i + \beta_2 F_i X_i + \beta_0 M_i + \beta_1 M_i S_i + \beta_2 M_i X_i + u_i \\ &= \alpha_0 F_i + \beta_1 (F_i S_i + M_i S_i) + \beta_2 (F_i X_i + M_i X_i) + \beta_0 M_i + u_i \\ &= \alpha_0 F_i + \beta_0 M_i + \beta_1 (F_i + M_i) S_i + \beta_2 (F_i + M_i) X_i + u_i \\ &= \alpha_0 F_i + \beta_0 M_i + \beta_1 S_i + \beta_2 X_i + u_i \qquad \text{since } F_i + M_i = 1 \ \forall\, i. \end{aligned}$$

$$\ln W_i = \alpha_0 F_i + \beta_0 M_i + \beta_1 S_i + \beta_2 X_i + u_i \qquad \mathbf{(13.0)}$$

- **In equation (12.1) – *females* as base group: set $\gamma_1 = 0$ *and* $\gamma_2 = 0$**

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + \gamma_0 M_i + \gamma_1 M_i S_i + \gamma_2 M_i X_i + u_i \qquad (12.1)$$

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + \gamma_0 M_i + u_i \qquad \mathbf{(13.1)}$$

- **In equation (12.2) – *males* as base group: set $\delta_1 = 0$ *and* $\delta_2 = 0$**

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \delta_0 F_i + \delta_1 F_i S_i + \delta_2 F_i X_i + u_i \qquad (12.2)$$

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \delta_0 F_i + u_i \qquad \mathbf{(13.2)}$$

- OLS estimation of any one of the three restricted regression equations (13.0), (13.1) or (13.2) on the pooled sample of 534 male and female employees yields the following value for the ***restricted*** **RSS**:

RSS_0 = 108.4501 *with* **df_0 = N – K_0 = 534 – 4 = 530**.

<u>The F-Test for Male-Female Equality of *Both* Slope Coefficients</u>

$$F = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N - K)} \sim F[df_0 - df_1, df_1]$$

$RSS_0 = 108.4501$ *with* $df_0 = N - K_0 = 534 - 4 = 530$
$RSS_1 = 106.4340$ *with* $df_1 = N - K = 534 - 6 = 528$

- ***Sample value* of the F-statistic:**

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$
$$= \frac{(108.4501 - 106.4340)/(530 - 528)}{106.4340/528} \qquad \Rightarrow \qquad \mathbf{F_0 = 5.001}$$
$$= \frac{2.0161/2}{106.4340/528}$$
$$= 5.001$$

- ***Null distribution* of F_0:** $F_0 \sim F[df_0 - df_1, df_1]$ = F[2, 528] under H_0.

- ***Decision Rule:*** At the 100α percent significance level

  1. ***reject* H_0** if $F_0 \ge F_\alpha[df_0 - df_1, df_1]$ *or* p-value for $F_0 \le \alpha$;
  2. ***retain* H_0** if $F_0 < F_\alpha[df_0 - df_1, df_1]$ *or* p-value for $F_0 > \alpha$.

- ***Critical Values of F[2, 528]:*** at the 5% and 1% significance levels.

  At α = 0.05: $F_{0.05}[2, 528] = 3.013$
  At α = 0.01: $F_{0.01}[2, 528] = 4.646$

- ***P-value for F_0* = 0.00706**.

- ***Inference:***

  Since $\mathbf{F_0 = 5.001 > 4.646 = F_{0.01}[2, 528]}$, ***reject* H_0** at the **1%** significance level.

  Since **p-value for F_0 = 0.00706 < 0.01**, ***reject* H_0** at the **1%** significance level.

4. Conditional Log-Wage Differentials Between Males and Females

Question: What is the **mean log-wage differential** between **male and female employees** with **the *same* education and work experience** – i.e., with **the *same* values of S and X?**

- **Conditional *female-male* mean log-wage differential** is:

$$E(\ln W \mid S_i, X_i, F_i = 1) - E(\ln W \mid S_i, X_i, F_i = 0)$$

or

$$E(\ln W \mid S_i, X_i, M_i = 0) - E(\ln W \mid S_i, X_i, M_i = 1).$$

- The ***female*** **log-wage equation** is

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i + u_i \qquad i = 1, ..., N_f = 245 \qquad \textbf{(2.1)}$$

Hence, the **conditional mean log-wage of *females* with S_i years of completed schooling and X_i years of work experience** is:

$$E(\ln W \mid S_i, X_i, F_i = 1) = E(\ln W \mid S_i, X_i, M_i = 0) = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i \qquad \textbf{(14.1)}$$

- The ***male*** **log-wage equation** is

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + u_i \qquad i = 1, ..., N_m = 289 \qquad \textbf{(2.2)}$$

Hence, the **conditional mean log-wage of *males* with S_i years of completed schooling and X_i years of work experience** is:

$$E(\ln W \mid S_i, X_i, F_i = 0) = E(\ln W \mid S_i, X_i, M_i = 1) = \beta_0 + \beta_1 S_i + \beta_2 X_i \qquad \textbf{(14.2)}$$

- The **conditional *female-male* mean log-wage differential** is obtained by ***subtracting* (14.2) from (14.1)**:

$$E\left(\ln W \middle| S_i, X_i, F_i = 1\right) = E\left(\ln W \middle| S_i, X_i, M_i = 0\right) = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i \quad \textbf{(14.1)}$$

$$E\left(\ln W \middle| S_i, X_i, F_i = 0\right) = E\left(\ln W \middle| S_i, X_i, M_i = 1\right) = \beta_0 + \beta_1 S_i + \beta_2 X_i \quad \textbf{(14.2)}$$

$$\begin{aligned} & E\left(\ln W \middle| S_i, X_i, F_i = 1\right) - E\left(\ln W \middle| S_i, X_i, F_i = 0\right) \\ &\quad = \alpha_0 + \alpha_1 S_i + \alpha_2 X_i - \beta_0 - \beta_1 S_i - \beta_2 X_i \\ &\quad = (\alpha_0 - \beta_0) + (\alpha_1 - \beta_1) S_i + (\alpha_2 - \beta_2) X_i \\ &\quad = \delta_0 + \delta_1 S_i + \delta_2 X_i \end{aligned}$$

where $\delta_0 = \alpha_0 - \beta_0$; $\delta_1 = \alpha_1 - \beta_1$; $\delta_2 = \alpha_2 - \beta_2$.

- **Pooled full interaction log-wage equation with *males* as base group**

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \delta_0 F_i + \delta_1 F_i S_i + \delta_2 F_i X_i + u_i \quad \textbf{(12.2)}$$

where $\delta_0 = \alpha_0 - \beta_0$; $\delta_1 = \alpha_1 - \beta_1$; $\delta_2 = \alpha_2 - \beta_2$.

- ***OLS estimation* of *pooled full interaction* log-wage equation (12.2) yields the sample regression equation**

$$\ln W_i = \hat{\beta}_0 + \hat{\beta}_1 S_i + \hat{\beta}_2 X_i + \hat{\delta}_0 F_i + \hat{\delta}_1 F_i S_i + \hat{\delta}_2 F_i X_i + \hat{u}_i \quad \textbf{(12.2*)}$$

The **OLS coefficient estimates (and t-ratios)** for (12.2) are:

$\hat{\beta}_0 = 0.6965$ (4.478)	$\hat{\beta}_1 = 0.09030$ (8.822)	$\hat{\beta}_2 = 0.01635$ (7.002)
$\hat{\delta}_0 = -0.3933$ (–1.605)	$\hat{\delta}_1 = 0.02144$ (1.318)	$\hat{\delta}_2 = -0.007491$ (–2.223)

♦ **The *estimate* of the conditional *female-male* mean log-wage differential is**

$$\hat{E}\left(\ln W \middle| S_i, X_i, F_i = 1\right) - \hat{E}\left(\ln W \middle| S_i, X_i, F_i = 0\right)$$

$$= (\hat{\alpha}_0 - \hat{\beta}_0) + (\hat{\alpha}_1 - \hat{\beta}_1)S_i + (\hat{\alpha}_2 - \hat{\beta}_2)X_i$$

$$= \hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i$$

$$\mathbf{= -0.3933 + 0.02144S_i - 0.007491X_i} \qquad \mathbf{(15)}$$

Example: The estimated **conditional *female-male* mean log-wage differential** for employees with **16 years of schooling and 10 years of work experience** is obtained by setting $\mathbf{S_i = 16}$ and $\mathbf{X_i = 10}$ in equation (15):

$$\hat{E}\left(\ln W \middle| S_i = 16, X_i = 10, F_i = 1\right) - \hat{E}\left(\ln W \middle| S_i = 16, X_i = 10, F_i = 0\right)$$

$$= -0.3933 + 0.02144 S_i - 0.007491 X_i$$

$$= -0.3933 + 0.02144(16) - 0.007491(10) = \mathbf{-0.12517}$$

Interpretation: The **average wage of *female* employees** with 16 years of schooling and 10 years of work experience is **approximately 12.5 percent *less*** than the **average wage of *male* employees** with the same schooling and work experience.

❑ **The *variance* of the conditional *female-male* mean log-wage differential** is

$$\text{Var}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \text{Var}(\hat{\delta}_0) + S_i^2 \text{Var}(\hat{\delta}_1) + X_i^2 \text{Var}(\hat{\delta}_2)$$
$$+ 2S_i \text{Cov}(\hat{\delta}_0, \hat{\delta}_1) + 2X_i \text{Cov}(\hat{\delta}_0, \hat{\delta}_2) + 2S_i X_i \text{Cov}(\hat{\delta}_1, \hat{\delta}_2)$$

The *standard error* of the conditional *female-male* mean log-wage differential is simply the ***square root* of the variance**:

$$\text{se}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \sqrt{\text{Var}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right)}$$

***Proposition*: The conditional *female-male* mean log-wage differential for employees with *16* years of schooling and *10* years of work experience *equals zero*.**

Null and Alternative Hypotheses

$$H_0\text{: } E\left(\ln W \mid S_i = 16, X_i = 10, F_i = 1\right) - E\left(\ln W \mid S_i = 16, X_i = 10, F_i = 0\right) = 0$$

or $$\delta_0 + \delta_1 S_i + \delta_2 X_i = \delta_0 + 16\,\delta_1 + 10\,\delta_2 = 0$$

$$H_1\text{: } E\left(\ln W \mid S_i = 16, X_i = 10, F_i = 1\right) - E\left(\ln W \mid S_i = 16, X_i = 10, F_i = 0\right) \neq 0$$

or $$\delta_0 + \delta_1 S_i + \delta_2 X_i = \delta_0 + 16\,\delta_1 + 10\,\delta_2 \neq 0$$

Perform a *two-tail* t-test: The ***t-statistic*** is

$$t\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \frac{\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i - (\delta_0 + \delta_1 S_i + \delta_2 X_i)}{\hat{se}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right)} \sim t[N - 2K_0] \qquad \textbf{(16)}$$

- **Calculate the *estimated variance* and *standard error*** of $\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i = \hat{\delta}_0 + 16\,\hat{\delta}_1 + 10\,\hat{\delta}_2$:

$$\hat{Var}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \hat{Var}(\hat{\delta}_0) + S_i^2 \hat{Var}(\hat{\delta}_1) + X_i^2 \hat{Var}(\hat{\delta}_2)$$
$$+ 2S_i \hat{Cov}(\hat{\delta}_0, \hat{\delta}_1) + 2X_i \hat{Cov}(\hat{\delta}_0, \hat{\delta}_2) + 2S_i X_i \hat{Cov}(\hat{\delta}_1, \hat{\delta}_2)$$

```
. regress lnw s x f fs fx

  Source |       SS       df       MS              Number of obs =     534
---------+------------------------------           F(  5,   528) =   41.68
   Model |  42.0079479     5  8.40158958           Prob > F      =  0.0000
Residual |  106.433962   528  .201579473           R-squared     =  0.2830
---------+------------------------------           Adj R-squared =  0.2762
   Total |   148.44191   533  .278502645           Root MSE      =  .44898

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       s |   .0903006   .0102354      8.822   0.000       .0701936    .1104076
       x |   .0163455   .0023344      7.002   0.000       .0117597    .0209314
       f |  -.3933286   .2450724     -1.605   0.109      -.8747652    .0881081
      fs |   .0214448   .0162689      1.318   0.188      -.0105149    .0534046
      fx |  -.0074912   .0033699     -2.223   0.027      -.0141112   -.0008712
   _cons |   .6964687    .155538      4.478   0.000       .3909194    1.002018
------------------------------------------------------------------------------
```

```
. matrix list VC2

symmetric VC2[6,6]
                s           x           f          fs          fx       _cons
    s   .00010476
    x   8.540e-06   5.449e-06
    f   .00151678   .00020904   .06006048
   fs  -.00010476  -8.540e-06  -.00381452   .00026468
   fx  -8.540e-06  -5.449e-06  -.00046734   .00001939   .00001136
_cons  -.00151678  -.00020904  -.02419208   .00151678   .00020904   .02419208
```

$$\hat{Var}(\hat{\delta}_0) = 0.06006048$$
$$\hat{Var}(\hat{\delta}_1) = 0.00026468$$
$$\hat{Var}(\hat{\delta}_2) = 0.00001136$$
$$\hat{Cov}(\hat{\delta}_0, \hat{\delta}_1) = -0.00381452$$
$$\hat{Cov}(\hat{\delta}_0, \hat{\delta}_2) = -0.00046734$$
$$\hat{Cov}(\hat{\delta}_1, \hat{\delta}_2) = 0.00001939$$

- **Set S_i = 16 and X_i = 10** in formula for $\hat{Var}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right)$:

$$\hat{Var}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \hat{Var}(\hat{\delta}_0) + S_i^2 \hat{Var}(\hat{\delta}_1) + X_i^2 \hat{Var}(\hat{\delta}_2) + 2S_i \hat{Cov}(\hat{\delta}_0, \hat{\delta}_1) + 2X_i \hat{Cov}(\hat{\delta}_0, \hat{\delta}_2) + 2S_i X_i \hat{Cov}(\hat{\delta}_1, \hat{\delta}_2)$$

- **Calculate *estimated variance* of** $\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i = \hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2$:

$$\begin{aligned}\hat{Var}\left(\hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2\right) &= \hat{Var}(\hat{\delta}_0) + (16)^2 \hat{Var}(\hat{\delta}_1) + (10)^2 \hat{Var}(\hat{\delta}_2) \\ &\quad + 2(16)\hat{Cov}(\hat{\delta}_0, \hat{\delta}_1) + 2(10)\hat{Cov}(\hat{\delta}_0, \hat{\delta}_2) + 2(16)(10)\hat{Cov}(\hat{\delta}_1, \hat{\delta}_2) \\ &= 0.06006048 + (16)^2(0.00026468) + (10)^2(0.00001136) \\ &\quad + 2(16)(-0.00381452) + 2(10)(-0.00046734) + 2(16)(10)(0.00001939)\end{aligned}$$

Results:

$$\hat{Var}\left(\hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2\right) = \mathbf{0.00374538}$$

$$\hat{se}\left(\hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2\right) = \sqrt{\hat{Var}\left(\hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2\right)} = \sqrt{0.00374538} = \mathbf{0.0611995}$$

- The ***sample value* of the t-statistic** under H_0 is calculated from (16) by setting

$$\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i = \hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2 = \mathbf{-0.12512354}$$

$$\delta_0 + \delta_1 S_i + \delta_2 X_i = \delta_0 + 16\delta_1 + 10\delta_2 = \mathbf{0}$$

$$\hat{se}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \hat{se}\left(\hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2\right) = \mathbf{0.0611995}$$

$$t\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right) = \frac{\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i - (\delta_0 + \delta_1 S_i + \delta_2 X_i)}{\hat{se}\left(\hat{\delta}_0 + \hat{\delta}_1 S_i + \hat{\delta}_2 X_i\right)} \qquad \mathbf{(16)}$$

$$t_0\left(\hat{\delta}_0 + 16\hat{\delta}_1 + 10\hat{\delta}_2\right) = \frac{-0.12512354 - 0}{0.0611995} = \frac{-0.12512354}{0.0611995} = \mathbf{-2.0445}$$

- ***Null distribution* of t_0:** $t_0 \sim t[N - 2K_0] = t[534 - 2(3)] = t[534 - 6] = t[528]$

- ***Two-tail critical values* of t[528]:** at the 5% and 1% significance levels

  $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$: $t_{0.025}[528] = 1.964$
  $\alpha = 0.01 \Rightarrow \alpha/2 = 0.005$: $t_{0.005}[528] = 2.585$

- ***Two-tail p-value* for t_0 = 0.04140**.

- ***<u>Inference</u>:***

  Since **$|t_0|$ = 2.045 < 2.585 = $t_{0.005}$[528]**, ***retain*** **H_0** at the ***1%* significance level**.
  Since **$|t_0|$ = 2.045 > 1.964 = $t_{0.025}$[528]**, ***reject*** **H_0** at the ***5%* significance level**.

  Since **p-value for t_0 = 0.04140 > 0.01**, ***retain*** **H_0** at the ***1%* significance level**.
  Since **p-value for t_0 = 0.04140 < 0.05**, ***reject*** **H_0** at the ***5%* significance level**.

- How to use *Stata* to compute this t-test: Use the following ***lincom*** command.

```
. lincom _b[f] + 16*_b[fs] + 10*_b[fx]

 ( 1)  f + 16.0 fs + 10.0 fx = 0.0

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     (1) |  -.1251235   .0611995     -2.045   0.041       -.245348   -.0048991
---------+--------------------------------------------------------------------
```